

"M-seeds" or seeds of meaning."

The process of identifying "meaningful N-word groups" is further described at paragraph 99 of the Specification, as step "e) Finding meaning conveying structures":

"To find meaning, means to determine what is the object (subject) that acts, how does it act (what does it do), upon which object (subject) does it act, and what is the result of the action. Other cases of finding out the meaning are related to the cases when we determine a) what are the properties of an object, b) what are the capabilities of the object, c) what are the results of action, and so on."

As further noted at paragraph 135, N-word groups are validated as M-seeds based upon an evaluation of their quantitative "significance" relative to other N-word groups identified during the windowing and scanning processes. Significance is defined in paragraph 25 as "Significance--a quantitative measure for each unit of text, the value of which can include the frequency of occurrence, the number of associative links, and the fitness to a user's needs."

As described in the Specification, the windowing and scanning operations identify several types of N-word groups, which may include a "before action part" (BAP) an "action part", and an "after action part". Identification of M-seeds is based upon determining the relative quantitative significance of such N-word groups. The Applicant submits that recitation of "meaningful N-word groups" in claims 1 and 18 is clearer than would be recitation of the term "M-seed" as used in the Specification, as the term "M-seed" is unique to the Specification, albeit well-defined therein.

Claims 2 and 9

Claims 2 and 9 stand rejected under 35 U.S.C. §112 as allegedly indefinite and lacking enablement on the basis of the Examiner's objection to the phrase "entity grouping rules". As described in paragraphs 139 to 150, the process of text generation proceeds "by combining relevant text units (text units that are marked by the pointers of the document trace) belonging to a requested level of generalization and satisfying constraints determined by the values of significance, or determined by a pre-assigned bias outlined by particular interests or goals of a user." Specific examples of rules determining how such text units are combined, are listed in paragraphs 142-148, and include using relationships among M-seeds to control the order by which sentences are constructed by merging text units containing similar before-action-parts (BAPs), after-action-parts (AAPs), and action-parts (APs). The strategies described in paragraphs 142-148 are generally referred to as "entity grouping rules" in paragraph 6 of the Specification, and consist of rules by which the analyzed structural entities of the input document are arranged to form the output text, as shown in FIG. 10. As can be seen by the more specific disclosure of paragraphs 142-148 relative to paragraph 6, the specific listing of rules later in the Specification is intended to define and enable the "entity grouping rules" recited generally in

paragraph 6, and in claims 2 and 9.

Rejections under 35 U.S.C. 102

Claims 1 and 18 have been rejected under 35 U.S.C. 102 on the basis of Mase et al., U.S. Pat. No. 5,978,820. In general, the Mase et al. patent describes a method of "summarization" that consists of a word-by-word analysis of text, in conjunction with some bulk statistics pertaining to the input document, in order to produce what might more accurately be deemed a condensation, rather than a summary based on meaning. In other words, the objective of the Mase et al. system is to filter out or eliminate insignificant portions of text, leaving a remainder of what is hypothesized to be the main point of the input text, based on identified types of sentences (see, e.g. Mase et al. Fig. 10, which shows a table of rules for extracting types of sentences from a newspaper article). This type of analysis is appropriate for Japanese text, of the type described in Mase, since Japanese text consists of characters which themselves convey larger units of meaning than English text. In other words, the meaning of a Japanese character is not as contextually dependent as that of an English word, since meaning is conveyed in English by groupings of words rather than single units. Additionally, there is no trading in Mase et al. that relates to the extraction of meaning on the basis of determined relatedness among conceptual elements of the input document.

In contrast, the Applicant's invention generally relates to extracting meaning from an input document by building a data structure that compares and categorizes text entities at multiple levels of granularity, instead of simply weighting individual words and sentences based on bulk statistics. The generation of output text is guided by the relationships among elements determined to convey meaning, rather than by a sequential elimination of text deemed to be statistically insignificant, as in the Mase et al. patent.

Turning to the Applicant's claims, the Examiner has applied Mase Figure 9; col. 6, lines 56-67; col. 7, lines 1-2; and col. 10 lines 27-54, as allegedly disclosing the Applicant's recited step of "determining text entities including meaningful N-word groups, phrases, simple sentences and compound sentences". In contrast, the referenced portion of Mase et al. discloses the computation of bulk statistics about the input document, such as the occurrence frequencies of types of sentences, and the identification of individual words which occur in the input document, but this portion of the Mase et al. patent does not describe the identification and determination of groups of words and phrases in order to determine meaning (as conveyed in written English). The cited passage of Mase et al. starting at column 6, line 56 refers to generating such bulk statistics in order to determine the "type" of text under analysis, but not as a process of categorizing text entities as particularly will be employed in Applicant's later recited steps of determining interconnections between these text entities for constructing a multigranular relational text structure.

The cited portion of Mase column 10 merely refers to the entire text of an input

document to be analyzed, and not to any grouping and categorization of text entities therein for the purpose of analyzing interconnections and constructing a relational structure at multiple levels of granularity for the purpose of extracting meaning. Furthermore, the Applicant's claimed method generates output text from the relational text structure (as noted, for example, in the entity grouping rules described at paragraphs 142 to 2148), which is independent of the narrative order of the input text. As can be seen by comparing the summaries generated by Mase (e.g. comparing the input/output text pairs of col. 10, lines 27-53 and col. 15, lines 15-25; and the "business letter" of col. 10 lines 60-67 with col. 15, lines 30-35) and the examples given in the Applicant's Specification, the Applicant's method of generating output text from a relational text structure built from a determination of the interconnections between text entities (in contrast to the Mase approach of simply statistically weighting individual sentences) produces an output text that is independent of the narrative order of the input text, yet preserves meaning in a manner that cannot be accomplished by applying a table of text elimination rules (as shown in Mase Fig. 10, for example).

Because the Mase et al. system does not disclose or suggest the construction of a relational structure based upon inter-relationships between identified units of meaning, phrases, simple and compound sentences, the Mase et al. system is limited to sequential filtering of text in order to eliminate portions deemed insignificant on the basis of a bulk statistical analysis of the entire document, and not an analysis based upon determination of meaningful text entities on a multigranular basis, as recited in Applicant's claims 1 and 18.

For the foregoing reasons, Applicant's claims 1 and 18 are believed to be patentably distinct relative to the Mase et al. disclosure. The remaining claims of the Application are dependent thereon, and are thus also believed to be allowable for at least the same reasons. Further distinctions of dependent claims 2 and 19, which serve as a basis for other dependent claims, are noted below.

Claims 2 and 19

As noted above, the Mase et al. text generation method proceeds on a sequential basis by eliminating text determined not to be significant. Claims 2 and 19 recite that text generation proceeds on the basis of entity grouping rules, such as are described at paragraphs 142-148 of the Specification. By constructing a relational text structure at multiple levels of granularity, as recited in respective claims 1 and 18, a set of such entity grouping rules can then be utilized to extract the output text from the structure, independent of the narrative structure of the input document. In rejecting claims 2 and 19, the Examiner refers to col. 9, lines 17-44 which describes "summarization rules" of the type employed by Mase et al. The Applicant submits that Mase's description of such mechanical "summarization rules", which proceed on a sequential basis through the input text, is distinct from the grouping rules claimed and described in the Application, which generates text on the extracting information by applying grouping rules to a relational structure built on the basis of meaning and

relationships determined throughout the input document.

Rejections under 35 U.S.C. §103

As noted above, claims 1 and 18 are believed to be distinct over the Mase et al. reference, which is the base reference for the proposed combinations of patents applied to the dependent claims. For at least the reasons cited above, the proposed combinations applied to the dependent claims fail to disclose or suggest the features of Applicant's dependent claims rejected under 25 U.S.C. §103, as the additional references fail to meet the deficiencies of the Mase et al. reference noted above.

The Kupiec et al. reference, applied to claims 3 and 20 describes a method of summarizing a document based solely on the selection or non-selection of sentences sequentially within a document. In that regard, the Kupiec et al. reference is cumulative to the disclosure of a similar concept in the Mase et al. reference, but does not meet the conditions of constructing a relational structure of the text entities recited in Applicant's claims 1 or 18, and then generating output text on the basis of extracting such entities from the relational structure. While claims 3 and 20 refer to utilizing the sequential order in the input text, this is stated as a user-selected feature, whereas in the Mase et al. and Kupiec et al. references, such operation is the only method possible, and is not established by entity grouping rules for extracting the output text from the database.

The Ozawa et al. reference, applied to claims 4-6, 14, 21-23, and 30, does not relate to text summarization at all, but describes a computer aided text generation system, which is provided as an authoring tool. Hence, it does not meet the deficiencies of the Mase et al. reference in providing the recited features of establishing entity grouping rules for extracting output text from a relational structure composed by multigranular analysis of an input document. Nor does the Mase et al. reference supply such teaching, as discussed above.

The Richardson et al. patent, applied to claims 9 and 26, describes a method for natural language parsing of sentences. Because the Richardson et al. patent is directed to parsing individual sentences, the Richardson et al. patent does not meet the deficiencies of the Mase et al. reference in constructing a relational structure of text entities determined from an input document as a whole for use in extracting a summary from such a relational structure in order to generate a summary for the document. As noted above, such elements are also missing from the Mase et al. reference.

The Tsourikov et al. reference, applied to claims 15, 16, 31, and 33, is similar to the Richardson et al. patent, in that it is directed to parsing individual sentences posed as natural language queries. The Tsourikov et al. reference describes comparing "subject-action-object" structures to similar structures within a database, but does not generate a relational database of such structures derived from multigranular analysis of a single input document for the purpose of then applying

entity grouping rules to produce a summary of the document. Hence, the proposed combination fails to disclose the features of the respective base claims 1 and 18, much less the additional features of the dependent claims.

The Hetherington patent, applied to claims 17 and 34, is directed to the object of producing a field-delimited database on the basis of input textual information, and is primarily concerned with categorizing elements of text within such free-format data records as addresses. The Hetherington patent is not addressed to the problem of determining relationships among text entities in a natural language input document, and similarly fails to meet the deficiencies noted above in connection with the Mase et al. patent.

In light of the foregoing, Applicant believes that this application is in form for allowance. The Examiner is encouraged to contact Applicant's undersigned attorney if the Examiner believes that issues remain regarding the allowability of this application.

Request for Reconsideration

For the foregoing reasons, the Applicant is believed to be distinct over the cited references and in condition for allowance. In the event that further discussions would lead to favorable reconsideration of the Applicant, the Examiner is encouraged to contact the undersigned attorney.

Respectfully submitted,

DANN, DORFMAN, HERRELL & SKILLMAN
A Professional Corporation
Attorneys for Applicant(s)

By


John B. Berryhill, Ph.D.
PTO Registration No. 36,452

Telephone: (215) 563-4100
Facsimile: (215) 563-4044